DOCKET NO.: HITACH 0018   PATENT 04 # 4/A

12 - 205-542

M.L.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re patent application of: ASHIDA, et al.

Serial No.:   09/994,951   Group No.: Unknown

Filed: November 27, 2001   Examiner: Unknown

For: **METHOD AND SYSTEM FOR DATABASE MANAGEMENT FOR DATA MINING**

RECEIVED

APR 2 9 2002

Technology Center 2100

Assistant Commissioner for Patents
Washington, D.C.   20231

PRELIMINARY AMENDMENT

Sir:

Please make the following change prior to examination of the above-referenced

application:

In the Specification:

Please amend the following:

Page 7, line 31, between "other" and "the," please insert --than--.

Page 9, line 3, between "the rule." and "A precision", please insert --200 people

satisfy the rule portion while 50 people satisfy both the rule and condition portions.--

A1

upon the selected customer lists 107 and the speculation models 110, the speculation

processing unit 111 generates speculation results 112.

5   Still referring to FIGURE 1, each of the above processing units processes

information in a predetermined sequence and manner. According to a predetermined rule

such as in an if-then format, the characteristic rule generation processing unit 103 extracts

certain characteristic information to generate the characteristic rules 104 based upon the

customer data 101, which includes at least one record each of which contains at least

record entries. After the characteristic rules 104 are generated by the characteristic rule

10   generation processing unit 103, the segment selection unit 106 determines the structure of

the multi-dimensional database based upon the data definition information 102. The

condition items in the data definition information 102 correspond to the key dimensions in

the multi-dimensional database while the conclusion items correspond to the analysis

dimensions. After the dimensional structure is determined, the characteristic rule

15   generation processing unit 103 loads the customer data 101 and generates the multi-

dimensional database. In other words, the above segment selection process includes two

types of tasks. One task is to generate multidimensional database using the condition items

as columns and rows, and the conclusion items as analysis results. The other task is to

output the selected customer list with the selected segment data in to the above created

20   multidimensional cells. A user is now involved to select one of the condition items in the

characteristic rules 104. In response to the above user selection, a display screen is

generated to display cell values as the conclusion items in the columns and rows which

specify the condition items.

25   One example of the customer data 101 is illustrated in FIGURE 2. The

exemplary customer data 101 is generally organized by the month, including March, April

and May. Within each month, the first column is a customer number or ID to identify a

customer, and for each identified customer, a record including information on

predetermined items such as gender, age, profit amount and cancellation status. Within

30   March, the cancellation status reflects an event between the beginning and the end of

March. On the other hand, information other than the cancellation status for the March

records is based upon the information at the end of January. For example, the customer

is between twenty and twenty-four and the gender is female, license is cancelled. A
rule/condition in the third column is a ratio between a number of records to satisfy the rule
and a number of records to satisfy only the condition portion of the rule. 200 people
satisfy the rule portion while 50 people satisfy both the rule and condition portions. A

5   precision level in the fourth column is a ratio between the number of records satisfying the
rule and the number of records satisfying the condition.

Now referring to FIGURE 5, an exemplary multidimensional display is
illustrated. In this example, the above rule No. 1 is selected in FIGURE 4. The selected

10   rule is that if the age is between twenty and twenty-four and the gender is female, license is
cancelled. Based upon the above selected rule, a multidimensional display screen displays
condition items as well as conclusion items, and the multidimensional display includes
rows for displaying age groups and columns for displaying gender. In each cell, the above
ratio between the number of cancelled customers for the rule and a total number of

15   customer is displayed as a conclusion item. The above ratio value is automatically
calculated by the system according to the current invention. The cells that meet the
conditions used in the selected rule are in a certain predetermined color in order to
distinguish at a first glance from other conditions that are not used in the rule. Other
conditions are displayed as pages of the multidimensional database.

20
Still referring to FIGURE 5, the display is modifiable. A user compares the cell
values of particular interest under the selected conditions to other cell values in order to
determine the validity or significance of the selected rule. Furthermore, the user constructs
other displays or speculation models and selects a segment to be used for the speculation

25   models by observing cell value changes after adding and deleting the conditions. The
addition and deletion of the conditions are generally based upon the user's opinion and
experience or even upon trials and errors. The conditions are changed by multi-
dimensional database functions such as drill up, drill down, slice and dice. In adding a
condition, one way is to drill down a page of the multi-dimensional database and to select a

30   slice. In deleting a condition, either a column or a row of a page in the multi-dimensional
database is drilled up. For example, the user moves a pointing device such as a mouse on a
triangle or an area indicating "ALL" in the profit amount and clicks the right mouse button

Within the function menu, the user selects a desired function by the left mouse button. Assuming that the user selects the selected customer list generation in the function menu and the March data is currently being displayed, the selected customer list 107 is selected from the customer data 101 from May or two months after the current data and only from a

5   portion that satisfies the imposed conditions 108. The month for the above analysis is automatically selected to be two months after the currently selected month. As described above with respect to FIGURE 2, certain portions of the data other than a specified data such as the cancellation status are automatically taken from two-months earlier. Next, assuming that the user selects the speculation mode generation in the function menu, the

10   speculation model generation unit 109 automatically generates an optimal speculation model based upon the conditions that the user has selected for the above described segment selection process or unit 106. Lastly, assuming that the user selects the speculation in the function menu, the speculation processing unit 111 automatically ~~conclude~~ concludes the speculation results 112 based upon the selected customer list 107 and the speculation

15   models 110. The speculation algorithm is substantially the same as the algorithm used for speculating the potential cancelled customers or possibility for the cancelled customers. The speculation algorithms include the prior art techniques that have been disclosed in the background section of the current application. The speculation item in the function menu remains disabled until the selected customer list 107 and the speculation models 110 have

20   been selected and successfully completed.

Now referring to FIGURE 7, a flow chart illustrates steps involved in a preferred process of the speculation model generation/selection according to the current invention. The steps are described with respect to the units and the data as shown in FIGURE 1. In a

25   step 701, a portion of the customer data 101 is selected according to the data definition information 102. In the step 701, the selected portion is further refined to extract records that satisfy the conditions as set forth in the selected segments 108. In a step 702, the extracted records in the step 701 are divided into model candidate data and validating data. For example, the division is accomplished by randomly sampling sixty percent of the

30   records as the model candidate data while the remaining forty percent as the validation data. After the division in the step 702, the conditions as defined in the data definition information 102 are comprehensively combined to generate in combination with the

Page 11, line 4, between "108." and "Next," please insert - -The month for the above analysis is automatically selected to be two months after the currently selected month. As described above with respect to FIGURE 2, certain portions of the data other than a specified data such as the cancellation status are automatically taken from two months earlier.- -

Page 11, line 9, please change "conclude" to - -"concludes".- -

Respectfully submitted,

Ken T. Yoshida, Esq.
Reg. No. 37,009

Date: December 4, 2001

KNOBLE & YOSHIDA LLC
Eight Penn Center, Suite 1350
1628 John F. Kennedy Blvd.
Philadelphia, PA 19103
(215) 599-0600